WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

FACE EMOTION BASED MUSIC RECOMMENDATION SYSTEM

Amey Pawar, Tanmay Kabade, Prasad Bandgar, Richa Chirayil, Prof. Tushar Waykole

(CS ,PCET's Nutan Maharashtra Institute of Engineering and Technology, Savitribai Phule Pune University, India)
(CS, NMIET,SPPU,Pune, India)

ABSTRACT

Nowadays, people tend to increasingly have more stress because of the bad economy, high living expenses, etc. Listening to music is a key activity that assists to reduce stress. However, it may be unhelpful if the music does not suit the current emotion of the listener. Moreover, there is no music player which is able to select songs based on the user emotion. Emotion detection is the process of detecting a human being's emotions based on various facial cues and visual information. This field has gained much traction since the popularity of deep learning. Emotion detection has also given rise to many applications that had not been thought of before. One of the areas that are heavily associated with emotions is music.

To solve this problem, this paper proposes an emotion-based music player, which is able to suggest songs based on the user's emotions; sad, happy, neutral and angry. This paper proposes a smart agent that sorts music according to the emotions expressed in each song and then suggests a playlist to the user based on his or her current mood. The user's local music collection is initially categorized based on the album's emotional impact. When a user wants to make a playlist based on their mood, they can do so at any time. At that specific moment, they take a snapshot of themselves. On this image, facial detection and emotion recognition techniques are used to recognize the user's emotion. The user is then presented with a music playlist that best matches this emotion. The main aim of this paper is to develop a low-cost music player that generates a sentiment-aware playlist automatically based on the user's emotional state. The software is designed to use as few computer resources as possible. The emotion module determines the user's emotion. The music classification module takes a record and extracts specific and important audio data.

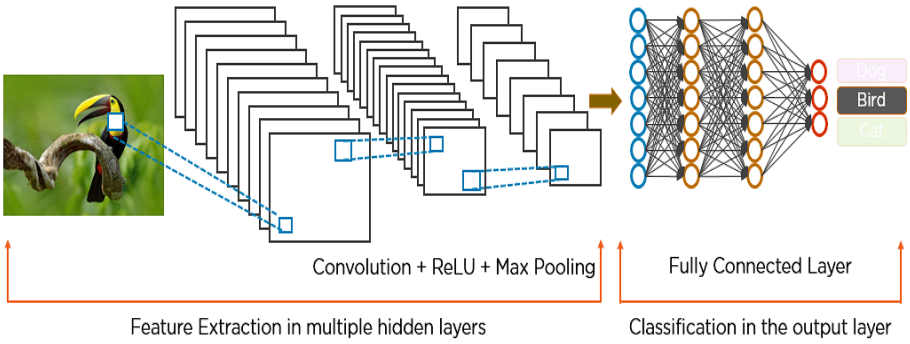
Keywords *– face emotion, pre-processing, classifier algorithm, feature extraction (CNN), open cv etc*

1. OBJECTIVE

Emotion recognition is a branch of artificial intelligence that is becoming increasingly important for automating processes that are inherently more time-consuming to complete manually.

Identifying an individual's mental state based on their feelings is a vital part of making efficient automatic decisions that are ideally suited to the person in question and can be useful in a number of situations.

2. SYSTEM ARCHITECTURE



A Convolutional Neural Network, also known as CNN or ConvNet, is a class of neural networks that specializes in processing data that has a grid-like topology, such as an image. A digital image is a binary representation of visual data. It contains a series of pixels arranged in a grid-like fashion that contains pixel values to denote how bright and what color each pixel should be.

A CNN typically has four layers: a convolutional layer, a Relu layer, a pooling layer, and a fully connected layer. The convolution layer is the core building block of the CNN. It carries the main portion of the network's computational load. This layer performs a dot product between two matrices, where one matrix is the set of learnable parameters otherwise known as a kernel, and the other matrix is the restricted portion of the receptive field. ReLU stands for the rectified linear unit. Once the feature maps are extracted, the next step is to move them to a ReLU layer.

ReLU performs an element-wise operation and sets all the negative pixels to 0. It introduces non-linearity to the network, and the generated output is a rectified feature map. The pooling layer replaces the output of the network at certain locations by deriving a summary statistic of the nearby outputs. This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights. The pooling operation is processed on every slice of the representation individually. There are several pooling functions such as the average of the rectangular neighborhood, L2 norm of the rectangular neighborhood, and a weighted average based on the distance from the central pixel. However, the most popular process is max pooling, which reports the maximum output from the neighborhood. In fully connected layer neurons have full connectivity with all neurons in the preceding and succeeding layer as seen in regular FCNN. This is why it can be computed as usual by a matrix multiplication followed by a bias effect. The FC layer helps to map the representation between the input and the output.

First the input image is given. Then the pixels of the image are feed to the convolutional layer that performs the convolution operation. This results in a convolved map. Then the convolved map is applied to a ReLU function to generate a rectified feature map. Then the image is processed with multiple convolutions and ReLU layers for locating the features. Different pooling layers with various filters are used to identify specific parts of the image. The pooled feature map is flattened and fed to fully connected layer to get the final output.

3. IMPLEMENTATION

A] SVM Algorithm:

Support Vector Machine is a learning algorithm that is mostly used for classifying data. It uses support vectors which are the coordinates of the observation from each data. From the plotted coordinates, a hyper-plane can be drawn where it can separate two or more classes. Since the hyperplane can be drawn anywhere, always select the hyper-plane which can separate the different classes better. This can be improved by maximizing the distances between the nearest data point and the hyper-plane. This distance is called the Margin. The System's performance is better if the margin is larger.

B] Pre-processing:

Pre-Processing plays a key role in overall process. Pre-Processing stage enhances the quality of input image and locates data of interest by removing noise and enhancing the image . It extracts the repetitious details from the image. Pre-Processing also includes filtering and normalization of image wherein the image after both process would be of same size but in a tilted manner.

C] Feature Extraction:

The output from the face detection stage forms an input to the feature extraction stage.. To obtain real-time performance and to reduce time complexity, for the intent of expression recognition, only eyes and mouth are considered. The combination of two features is adequate to convey emotions accurately. Finally, In order to identify and segregate feature points on face, point detection algorithm is used.

- **Eye Extraction**: The eyes display strong vertical edges (horizontal transitions) due to its iris and eye white. In order to find the Y coordinate of the eyes, vertical edges from the horizontal projection of the image is obtained through the use of Sobel mask.

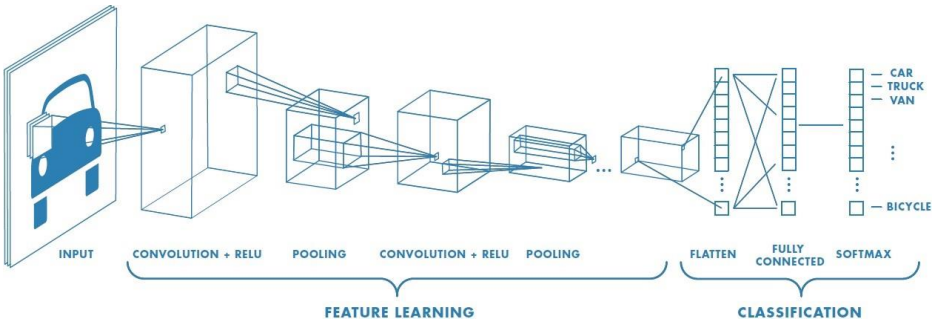
- **Eyebrow Extraction:** Two rectangular regions in the edge image which lies directly above each of the eye regions are selected as the eyebrow regions. The edge images of these two areas are obtained for further refinement. Now Sobel method was used in obtaining the edge image as more images can be detected when compared to Robert's method. These obtained edge images are then dilated and the holes are filled. The result edge images are used in refining the eyebrow regions.

- **Mouth Extraction**: The points in the top region, bottom region, right corner points and left corner points of the mouth are all extracted and the centroid of the mouth is calculated

D] CNN:

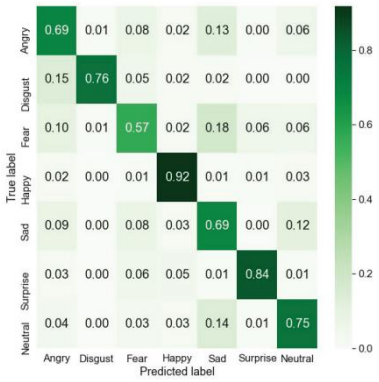
A Convolutional Neural Network, also known as CNN or ConvNet, is a class of neural networks that specializes in processing data that has a grid-like topology, such as an image. A digital image is a binary representation of visual data. It contains a series of pixels arranged in a grid-like fashion that contains pixel values to denote how bright and what color each pixel should be. The human brain processes a huge amount of information the second we see an image. Each neuron works in its own receptive field and is connected to other neurons in a way that they cover the entire visual field. Just as each neuron responds to stimuli only in the restricted region of the visual field called the receptive field in the biological vision system, each neuron in a CNN processes data only in its receptive field as well. The layers are arranged in such a way so that they detect simpler patterns first (lines, curves, etc.) and more complex patterns (faces, objects, etc.) further along. By using a CNN, one can enable sight to computers.

A CNN typically has three layers: a convolutional layer, a pooling layer, and a fully connected layer.



4. CONFUSION MATRIX

The confusion matrix of the Ensemble model is shown in fig. The rows correspond to the true values, and the columns correspond to our predictions. As we can clearly see, Fear is the class where our network fares the worst, and Happy is the most successful class. Another interesting observation is that 18 percent of the images labeled as fear are predicted as sad by our model, which is similar to human's mispredictions on the same image.



Confusion Matrix

5. WEB-APP

We have used the chrome browser application to run the Web-App. The flow of this application is as follows:

1) Run the capture.py file. It will then trigger the HTML file, which will show the CSS-HTML based music player (web-page)

2) To play any music, click on the play button shown on the song or have a plus sign to add it to the queue.

3) Another option based on emotion will be shown on the right upper side, select it. JavaScript will trigger the python function.

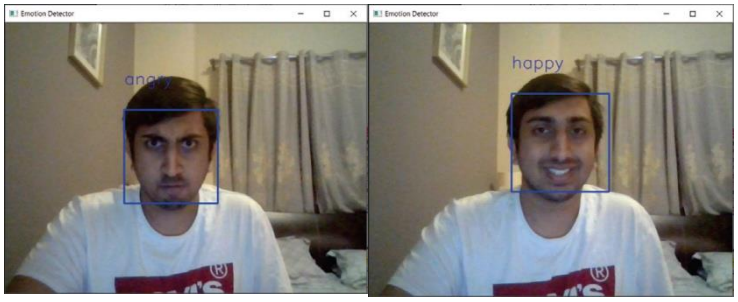
4) Camera will start and record the back-end image and go for ten successful images that contain any face.

5) Generate emotion prediction on those images, get the aggregate result of those ten results, choose appropriate emotion, and forward it to JS script.

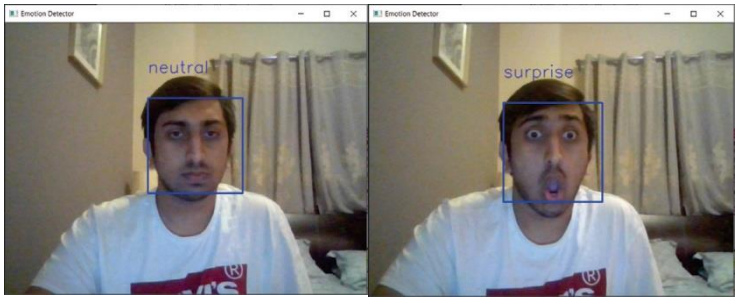
6) JS chooses a random song of that genre to play.

7) Whenever a song will go to an end, it will repeat the same back-end process so that the user will not be aware of it.

Fig shows the emotion detection processing action

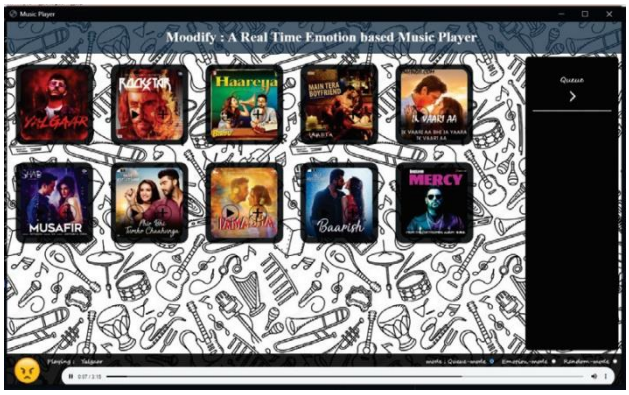


(a)Happy **(b)Angry**



(c)Neutral **(d)Neutral**

Fig shows the front-end of the web-app display



6. CONCLUSION

Music plays a major role in handling the stressful situations and emotions triggers of the user. So, it is required to recommend music that suits the current emotional needs of the user. There already exists widely used audio and video recommender systems like Spotify, Netflix, Gaana, etc which work based on search queries and not emotional needs of the user. So, the proposed CNN based model detects the emotion and generate the playlist accordingly. The model is embedded with modules for detecting facially expressed emotions. We have developed a model that plays a song based on the user emotions. The fundamental purpose of the system is to change or maintain the emotional state of the user and match personal music preferences by exploring music tracks with specific attributes. The Emotion-Based Music Player is used to facilitate the use by physically challenged people to automate and give them better music player experience.

REFERENCES

[1] B. Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information," Sensors (Basel, Switzerland), vol. 18, 2018.

[2] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. C. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. T. Ionescu, M. Popescu, C. Grozea, J.

Bergstra, J. Xie, L. Romaszko, B. Xu, C. Zhang, and Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," Neural networks : the official journal of the International Neural Network Society, vol. 64, pp. 59–63, 2015.

[3] Y. Tang, "Deep Learning using Linear Support Vector Machines," arXiv: Learning, 2013.

[4] L. Deng, "The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]," IEEE Signal Processing Magazine, vol. 29, pp. 141–142, 2012

[5] A. Krizhevsky, "CIFAR-10 (Canadian Institute for Advanced Research)." [Online]. Available: http://www.cs.toronto.edu/ kriz/cifar.html

[6] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," ArXiv, vol. abs/1804.08348, 2018.

[7] D. V. Sang, N. V. Dat, and D. P. Thuan, "Facial expression recognition using deep convolutional neural networks," 2017 9th International Conference on Knowledge and Systems Engineering (KSE), pp. 130– 135, 2017.

[8] C. Pramerdorfer and M. Kampel, "Facial Expression Recognition using Convolutional Neural Networks: State of the Art," ArXiv, vol. abs/1612.02903, 2016.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, pp. 84 – 90, 2012.

[10] Z. Yu and C. Zhang, "Image based Static Facial Expression Recognition with Multiple Deep Network Learning," Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, 2015.